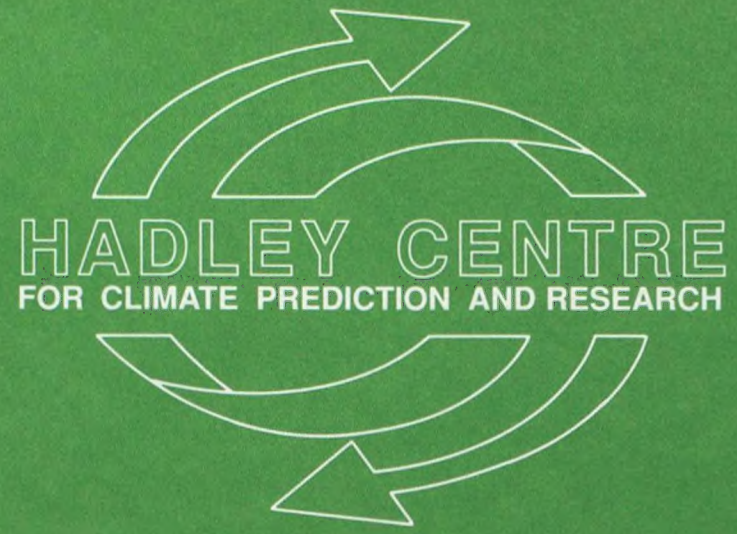


Development of worldwide marine data eigenvectors since 1985.

by

A. W. Colman

CRTN 28

June 1992

CLIMATE RESEARCH TECHNICAL NOTE

Hadley Centre
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SY

CLIMATE RESEARCH TECHNICAL NOTE NO. 28

DEVELOPMENT OF WORLDWIDE MARINE DATA EIGENVECTORS SINCE 1985

by

A W COLMAN

Hadley Centre for Climate Prediction and Research
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SY
U.K.

NOTE: This paper has not been published. Permission to quote from it should be obtained from the Director of the Hadley Centre.

DEVELOPMENT OF WORLDWIDE MARINE DATA EIGENVECTORS SINCE 1985

A.W.COLMAN MARCH 1992

ABSTRACT

This note documents the background to, and methods used, to develop global marine surface temperature anomaly eigenvectors (EOFs). It is not intended as a detailed note on the scientific results though some of these will be mentioned as appropriate.

Seasonal to century time-scale variability in marine temperature anomalies has been extensively investigated using the Meteorological Office historical sea surface (SST) and night marine air temperature (NMAT) datasets and a number of papers have been published making use of them. Eigenvector analysis is a proven technique of finding patterns of spatial variability and their importance, though deductions can only be soundly drawn with the help of independent evidence. Production of global marine temperature eigenvectors was begun in 1984. Since then there have been continual improvements to the quality control and analysis methods and in the completeness of the data. Limited work on combined ocean and land surface data will not be discussed here.

In preparation for eigenanalysis, quality controlled 5°x5° (latitude x longitude) monthly data are averaged to form three month seasonal means for larger areas (at least a 10°x10° area). Additional quality control procedures are used to remove excessively large anomalies and smooth excessive gradients. Corrections are made for uninsulated bucket SST measurements prior to 1942 and gaps are filled to produce a nominally complete dataset for eigenanalysis.

Many regional eigenvector patterns have been produced and used in papers. Here we confine discussion to either global patterns or regional patterns derived from an analysis of global data. Recognizable global patterns include one showing the spatial distribution of the century time scale global SST trend and a second showing El Nino related anomalies. A third global eigenvector showing a contrast between northern and southern hemisphere anomalies has been a useful tool in predicting Sahel rainfall and yet another has similarities to El Nino but seems to vary on bidecadal time scales.

1. INTRODUCTION

SST variations exert a strong influence on atmospheric climate because of the large heat capacity of the oceans relative to the atmosphere and the large spatial extent of the oceans. Sea temperature anomalies persist on longer timescales than most atmospheric variables and can have a surprising magnitude and spatial scale. Hence they are potentially good predictors of atmospheric variability on seasonal and perhaps longer timescales and may provide vital insights into decadal to century time scale variability. Because of the great usefulness of SST data in climate studies, the Hadley Centre has developed global datasets of SST and NMAT which have recently become probably the most complete available.

2.THE HADLEY CENTRE MARINE DATASETS

The MOHSST3 (Meteorological Office Historical Sea Surface Temperature) (Minhinick and Folland, 1984) and MOHMAT2 (Meteorological Office Historical Marine Air Temperature) datasets were completed in 1983. These datasets contain 5° latitude x 5° longitude monthly mean temperatures and anomalies from a 1951-80 average. MOHMAT2 has since been updated to 1989 and MOHSST3 and an updated version, MOHSST4, are discussed in Bottomley et al. (1990). Data sources are mainly ship and, recently, buoy observations from the Meteorological Office Marine Data Bank (Shearman, 1983). A further update to MOHSST (MOHSST5) has recently been completed (May 1991). To create MOHSST5, gaps in the MOHSST4 data were filled where possible with 5°x5° data from the Comprehensive Ocean Atmosphere Data Set (COADS) (Woodruff et al,1987). This combination improves upon MOHSST4 coverage by up to 50% of the earlier coverage for some late nineteenth century and early 20th century decades. The quality control scheme described in section 5 was then applied to remove outliers. Figure 1 shows the improvement in data coverage of MOHSST5 compared with MOHSST3 due to extra observations and the further improvement in coverage due to the interpolation scheme described in section 5.

MOHSST datasets are constructed from historical data stored in the Meteorological Office Marine Data Bank by a complex procedure whose essence is as follows (full details in Bottomley et al., 1990). Observations are averaged to form 1°x1° pentad means which are converted to anomalies by subtracting a background average, currently a set of 1951-80 normals calculated from Meteorological Office data and the Alexander and Mobley climatology (1976). The 1°x1° pentad anomalies in a given calendar month and 5°x5° area are first winsorized, though only if at least 4 1°x1° anomalies are being averaged. In this process the lowest 25% of anomalies are all reset to the lower quartile boundary and the highest 25% of anomalies are all reset to the upper quartile boundary. These data are then averaged to create 5°x5° anomalies. 1951-80 normals can be added to form 5°x5° monthly means. Finally, corrections are added to the pre-1942 SST anomalies to compensate for biases resulting from the extensive use of uninsulated sea buckets to collect the sea water and make SST measurements (Folland, 1991, gives full details of the models). MOHSST1-4 were made using solely this technique. For the preparation of MOHSST5, gaps in 5°x5° anomalies derived from MOHSST4 are filled with data from COADS, if available, with the same bucket corrections added. Subsequently the interpolation process described in section 5 is applied.

2.1 History of the marine datasets used for Eigenanalysis

These are summarized below:

1983-Sept 1986: MOHSST3 5°x5° monthly means for 1856-1981 and normals for 1951-80.
0.3°C added to pre-1942 data to correct for uninsulated bucket measurements (Folland, Parker and Kates, 1984).

Oct 1986-Oct 1987: MOHSST3 5°x5° monthly means for 1856-1981.
Normals calculated using MOHSST3 and MIT data for 1951-80.
Mk.1 uninsulated bucket model corrections (Folland, Parker and Chadwick, 1986, Folland and Hsiung, 1987) added to SST.

Nov 1987-Oct 1989: MOHSST3 5°x5° monthly means for 1856-1981. Same as above but with Mk.2 uninsulated bucket model corrections (GOSTA Atlas table 2B) added to SST and improved 1951-80 normals.

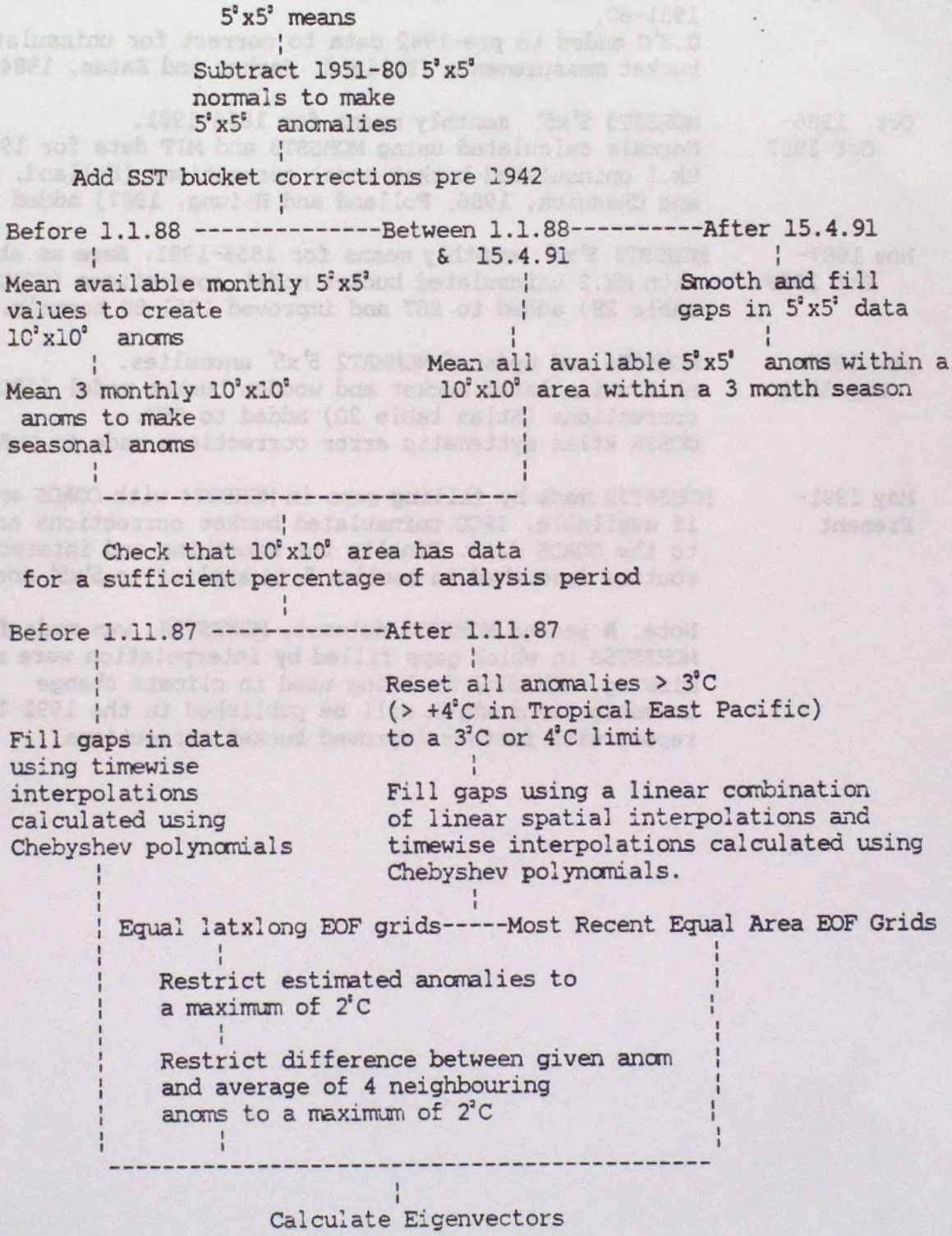
Nov 1989-May 1991: MOHSST4 and updated MOHMAT2 5°x5° anomalies.
Mk.3 uninsulated bucket and wooden bucket model (IPCC) corrections (Atlas table 2D) added to SST.
GOSTA Atlas systematic error corrections made to NMAT

May 1991-Present: MOHSST5B made by filling gaps in MOHSST4 with COADS anomalies if available. IPCC uninsulated bucket corrections are added to the COADS data. Finally the smoothing and interpolation routine described in section 5 is applied to 5°x5° anomalies.

Note. A second MOHSST5 dataset, MOHSST5A, was made from MOHSST5B in which gaps filled by interpolation were reset to missing. MOHSST5A is being used in climate change investigations which will be published in the 1992 IPCC report with further improved bucket corrections.

2.2 Preparation of the marine data for eigenanalysis

TABLE 1.FLOW DIAGRAM SUMMARIZING THE PREPARATION OF DATA FOR EIGENANALYSIS



* When coarser grids (10°x20° or 10°x30° or equal area) are used, read the coarser grid unit in place of 10°x10°. Otherwise data preparation is the same.

3. SYSTEMATIC ERROR CORRECTIONS INCLUDING CORRECTIONS FOR UNINSULATED BUCKET MEASUREMENTS

The observation that enhanced seasonal variations occur in pre-1942 SST anomalies, but not in more recent data, and the existence of a discontinuity in many SST data between 1941 and 1942 showed a requirement for systematic corrections to be made. The main cause of these systematic errors is believed to the measurements from uninsulated canvas buckets (Bottomley et al.,1990, and Folland and Parker, 1990) which were largely discontinued in favour of engine intake measurements in 1942 due to wartime conditions. The buckets were lowered over a ships side to collect the sea water then the SST was measured on deck by placing a thermometer in the retrieved bucket. Whilst waiting for the thermometer to adjust to the correct temperature, the sea water in the bucket would usually cool down, or occasionally warm up, depending on the humidity, wind speed, difference between the air and sea temperature and to a small extent, the solar radiation. (Folland, 1991)

In 1984 a constant addition of 0.3°C was made to all pre-1942 SST (Folland, Parker and Kates, 1984). Since 1986 models of the physics of uninsulated canvas buckets and more recently of semi-insulated wooden buckets have been used to calculate corrections dependent on month and 5°x5° location (Parker and Folland, 1990 and Folland, 1991) and on the size of the anomalous annual cycle in the data relative to 1951-80 cycles. These corrections take account of climatological mean weather conditions as indicated in the previous paragraph and also an assumed mean ship speed that changes with time. The corrections have evolved, e.g. those applied to MOHSST5 are slightly different from earlier versions. Smaller corrections are also made to the NMAT data (see Bottomley et al.,1990). Corrections for known truncation errors (data too cold) have also made to Japanese North Pacific SST data in the 1930s since Nov 1989.

4. QUALITY CONTROL

The problem is to balance the shortage of available observations from early years with the need for a complete dataset for eigenanalysis. Thus the choice of quality control flags has corresponded to fairly generous criteria. Seasonal data form the shortest unit time scale of analysis and 10°x10° data the smallest space scale. Thus for MOHSST3-based eigenvectors a seasonal 10°x20° mean anomaly may occasionally be based on as little as one 1°x1° pentad mean anomaly. For MOHSST4 and MOHSST5-based eigenvectors quality control was a little more strict; 5°x5° monthly anomalies require a minimum of two 1°x1° pentad anomalies.

The maximum absolute value allowed for all monthly MOHMAT and MOHSST 5°x5° anomalies is ±7.5°C, larger values being regarded as missing. This limit is not strict enough for average values over 10°x10° or larger size areas derived from the 5°x5° values (a 10°x10° or larger area seasonal anomaly may be based on only one 5°x5° monthly anomaly). Thus anomalies for 10°x10° or larger areas were limited to ±3°C except for tropical east Pacific (180°N-80°W,10°N-10°S) where positive anomalies up to +4°C are allowed. This 3°C limit may be too small for NMAT in high latitudes in winter and has now been raised to 4°C.

For any analysed area to be used in eigenanalysis, at least 60% of seasons must have data before interpolation. The exception is MOHSST5 data where the 5°x5° interpolation process described in section 5 is carried out before the 60% completeness criterion is applied.

5. SMOOTHING AND FILLING GAPS IN MOHSST5 5°x5° SEA TEMPERATURES

An additional quality control procedure was used with MOHSST5 data. This routine has been developed to improve the reliability and coverage of the 5°x5° data and has two main objectives:

A) To remove suspicious spikes (sudden large positive or negative values) in the anomaly time series for each area. One reason for the spikes is the fact that the quality control method used with the COADS data is less strict than that applied to Marine Data Bank data. However spikes were also present in MOHSST3 and MOHSST4-based analyses.

B) To modestly increase data coverage in as sound a way as possible so as to reduce the subsequent use of interpolation based on Chebyshev polynomials. The latter technique, as applied here, ignores the intrinsic variability on small timescales (1-3 months) and on small space scales (5°x5° resolution). We preserve additional information on these scales by prior interpolation of monthly data. Care needs to be taken not to interpolate from bad data, however.

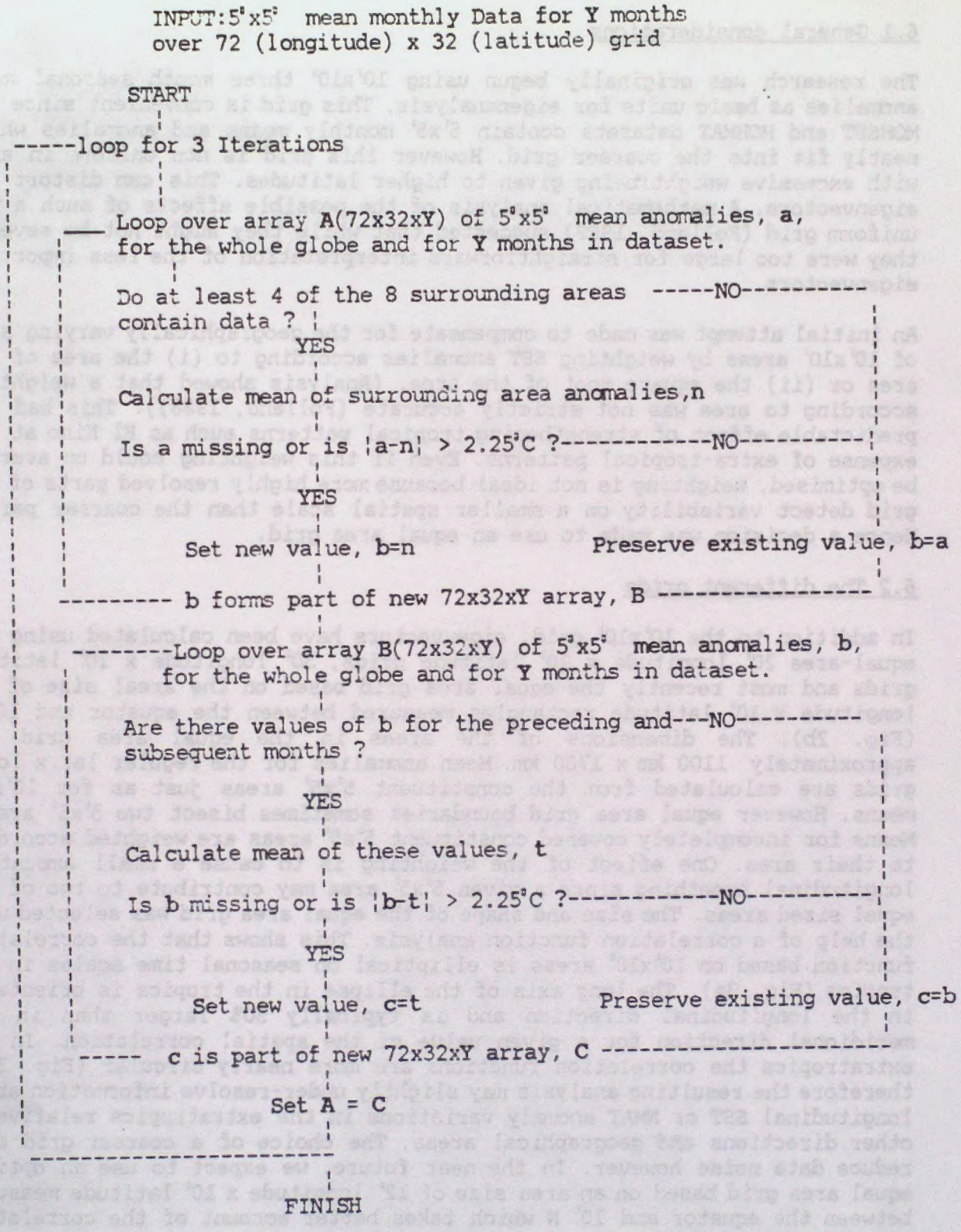
Table 2 is a flow diagram of the additional interpolation process used with MOHSST5. The process is one of simple linear interpolation based on adjacent values in space and time and operates on the monthly data. Firstly, gaps are filled or replacements made if there are four or more adjacent 5°x5° boxes containing data (out of a maximum of eight). The average of these box anomalies is used to fill a gap or replace a monthly data value if the value differs by more than 2.25°C from this average. When this process has been completed for every 5°x5° area, further gaps are filled or replacements made using the average anomaly for the months preceding and subsequent to the gap or outlier, provided there is data for both adjacent months. Further replacements are made if the average of the preceding and following value differs by 2.25°C from the value under test. Iteration sometimes reveals additional doubtful values; hence the whole process is repeated three times. For example, an individual outlier may influence a surrounding average so that values estimated using this surrounding average may need to be replaced.

The number of iterations (three) and the minimum number of surrounding 5°x5° areas that were required to have data to detect unlikely values in a given area and month were chosen by trial and error. Particular care went into the choice of the difference criterion. Tests were made with values of 1, 1.5, 2, 2.25 and 2.5°C while the minimum number of surrounding areas was also varied. The final judgement was made with the help of 14 monthly charts covering the annual cycle from the 1860s onwards.

To complete three iterations of the timewise interpolation procedure, data are required for the three months preceding or subsequent to the month being processed. This causes a problem at the beginning and end of the series. Thus a complete interpolated dataset must finish three months before the present. At the beginning of the dataset, the interpolation process is simplified. For the first month in the dataset, January 1856, only spatial interpolation is used. For February and March 1856, January data is used instead of older data.

This interpolation improves data coverage throughout the period represented by historical observations (Fig. 1) but the interpolation is most useful in filling gaps in the older data.

TABLE 2


INPUT:5°x5° mean monthly Data for Y months
over 72 (longitude) x 32 (latitude) grid
START
loop for 3 Iterations
Loop over array A(72x32xY) of 5°x5° mean anomalies, a,
for the whole globe and for Y months in dataset.
Do at least 4 of the 8 surrounding areas contain data ?
NO
YES
Calculate mean of surrounding area anomalies,n
Is a missing or is |a-n| > 2.25°C ?
NO
YES
Set new value, b=n
Preserve existing value, b=a
b forms part of new 72x32xY array, B
Loop over array B(72x32xY) of 5°x5° mean anomalies, b,
for the whole globe and for Y months in dataset.
Are there values of b for the preceding and subsequent months ?
NO
YES
Calculate mean of these values, t
Is b missing or is |b-t| > 2.25°C ?
NO
YES
Set new value, c=t
Preserve existing value, c=b
c is part of new 72x32xY array, C
Set A = C
FINISH
OUTPUT:5°x5° mean monthly Data for Y-6 months
over 72 (longitude) x 32 (latitude) grid

6.THE EIGENVECTOR GRID

6.1 General considerations

The research was originally begun using 10°x10° three month seasonal mean anomalies as basic units for eigenanalysis. This grid is convenient since the MOHSST and MOHMAT datasets contain 5°x5° monthly means and anomalies which neatly fit into the coarser grid. However this grid is non uniform in area with excessive weight being given to higher latitudes. This can distort the eigenvectors. A mathematical analysis of the possible effects of such a non uniform grid (Folland, 1988) suggested that while they might not be severe, they were too large for straightforward interpretation of the less important eigenvectors.

An initial attempt was made to compensate for the geographically varying size of 10°x10° areas by weighting SST anomalies according to (i) the area of the area or (ii) the square root of the area. (Analysis showed that a weighting according to area was not strictly accurate (Folland, 1988)). This had the predictable effect of strengthening tropical patterns such as El Nino at the expense of extra-tropical patterns. Even if this weighting could on average be optimised, weighting is not ideal because more highly resolved parts of the grid detect variability on a smaller spatial scale than the coarser parts. Hence a decision was made to use an equal area grid.

6.2 The different grids

In addition to the 10°x10° grid, eigenvectors have been calculated using non equal-area 20° longitude x 10° latitude grids, 30° longitude x 10° latitude grids and most recently the equal area grid based on the areal size of 15° longitude x 10° latitude rectangles measured between the equator and 10° N (Fig. 2b). The dimensions of the areas in the equal area grid are approximately 1100 km x 1700 km. Mean anomalies for the regular lat.x long. grids are calculated from the constituent 5°x5° areas just as for 10°x10° means. However equal area grid boundaries sometimes bisect two 5°x5° areas. Means for incompletely covered constituent 5°x5° areas are weighted according to their area. One effect of the weighting is to cause a small amount of longitudinal smoothing since a given 5°x5° area may contribute to two of the equal sized areas. The size and shape of the equal area grid was selected with the help of a correlation function analysis. This shows that the correlation function based on 10°x10° areas is elliptical on seasonal time scales in the tropics (Fig. 3a). The long axis of the ellipse in the tropics is orientated in the longitudinal direction and is typically 50% larger than in the meridional direction for a given value of the spatial correlation. In the extratropics the correlation functions are more nearly circular (Fig. 3b), therefore the resulting analysis may slightly under-resolve information about longitudinal SST or NMAT anomaly variations in the extratropics relative to other directions and geographical areas. The choice of a coarser grid does reduce data noise however. In the near future, we expect to use an optimal equal area grid based on an area size of 12° longitude x 10° latitude measured between the equator and 10° N which takes better account of the correlation function analyses.

7.FILLING DATA GAPS IN THE EIGENVECTOR GRID

Eigenanalysis requires a nominally complete dataset. Gaps are left in the eigenvector grid data even after the 5°x5° scale interpolation described in section 5 has been applied. Remaining gaps have been filled using a

combination of Chebyshev polynomial interpolations or extrapolations of the SST anomaly time series for each area and averages of anomalies for neighbouring areas. All existing anomalies are treated as real at this stage, even if they are based on the interpolation of 5°x5° area data described above. Interpolation is only applied to 3 month seasonal anomalies. Annual anomalies, also used to make eigenvectors, are calculated by averaging the 4 relevant seasonal anomalies, interpolated or uninterpolated.

7.1 Chebyshev Polynomial Interpolations

Initially this was the only technique used to fill data gaps. It was chosen because it can provide a smoothed model of the time series on many frequencies throughout the time series. Hence it is a good technique for filling large gaps without resorting to spatial interpolation or interpolations from related variables. High order terms in the Chebyshev series tend to decrease in magnitude more quickly than in alternative series such as power series hence the number of terms used can be kept low and computing time saved without loss of accuracy. Excluding spatial relationships from the interpolation prevents artificial spatial patterns developing. This is important as unreal periods of spatial coherence can corrupt the eigenvector patterns. Artificial timewise relationships are limited to individual areas and are incoherent on a large spatial scale and hence do not contribute to the high order EOFs.

The Chebyshev polynomial (Burden and Faires, 1989) time series of degree k is of the form:

$$f(x) = \tfrac{1}{2}a_0T_0(x) + a_1T_1(x) + a_2T_2(x) + \ldots + a_kT_k(x), \qquad \ldots(1)$$

where $T_n(x) = \cos(n \arccos x)$ and x is a time variable, here linearly related to season or year.

For filling gaps in SST or NMAT data, k has been set to 10. This value was chosen by trial and error by observing the results of using a number of different values of k and choosing a value which seemed to best estimate missing values.

The constants a are derived so as to on average produce a least squares fit to the observed data over the whole period analysed (typically between 30 and 90 years or between 120 and 360 three month seasons), ie. to minimize $\Sigma(f_x-y_x)^2$ where y_x is the observed SST anomaly for season x. The fit is done separately for each area.

Equation 1 requires that the absolute value of x is less than or equal to 1; hence x is derived from the time variable (season date) t using the transformation:-

$$x=\frac{2t-t(\max)-t(\min)}{t(\max)-t(\min)} \qquad \ldots(2)$$

where t(min) and t(max) refer to the first and last season respectively in the analysis period. Only data for consecutive seasons is interpolated; thus eigenvectors using annual data are based on an appropiate averaging of seasonal interpolated values. So for an analysis over 360 consecutive seasons, x would have values of -1,-357/359,-355/359,....,+1.

The Chebyshev polynomial interpolation technique used here does have a number of faults. Chebyshev polynomial curves of order 10 have considerably less variance than real data. This is partly deliberate so as not to fit the polynomials to noise. High frequency variability can be missed when gaps are large and is usually not regular enough over the whole eigenanalysis period (30-90 years) over which the polynomial is calculated to be modelled by the polynomial. This may artificially reduce the percentage of total variance explained by these eigenvectors which model high frequency variations and also distort their time series. However it is necessary to use a long period to adequately fill long gaps. Problems can nevertheless arise if much missing data is concentrated at a particular time (eg. early in the twentieth century or during the war years). Thus it is inevitable that if many consecutive data are missing, the magnitude of an important ENSO may be greatly under represented. This will tend not to happen if only a few isolated data are missing. Chebyshev polynomial curves also sometimes show rather wild high frequency oscillations particularly near their beginning or end if little real data is present. The effect of these wild oscillations have been restricted by setting strict limits ($\pm 2^{\circ}C$) to the magnitude of anomalies interpolated using this method. A more complex set of criteria could be used to select the polynomials but this was not tried. It is felt that more advanced techniques of interpolation are needed to make a major impact on this problem. However the results of this method seem to be satisfactory (see section 10).

7.2 Spatial Interpolations

A second, spatial, interpolation technique is used to create a second estimate of missing values in most cases (see below). The spatial interpolation only uses data from nearby areas to fill gaps, hence minimizing the risk of producing artificial spatial patterns. This is combined with the estimate from the Chebyshev time series method to combine, in a simple way, spatial and temporal information. Two slightly different schemes combining these timewise and spatial interpolation techniques have been used for regular Mercator and equal area grids respectively.

7.2.1 Spatial and Temporal Interpolation for data on a Regular Grid

The following shows how an estimate of a missing value is made for season t and area a. This interpolation is additional to that used to create the MOHSST5 dataset. This is calculated as a weighted average of (a) the Chebyshev polynomial estimate (b) the north/south and (c) the east/west linear interpolation estimates. If neighbouring data is complete then the polynomial and two linear interpolations are weighted in the ratios 50%:25%:25%. The north/south linear interpolation is calculated by averaging the anomalies from adjacent areas to the north and south and the east/west interpolation is calculated by averaging anomalies from adjacent areas to the east and west. If one or more of these adjacent area anomalies is missing, the relevant linear interpolation weight is set to zero. The polynomial weight is also reduced in the case of long time gaps in proportion to the reciprocal of the length of time gap. In the worst cases, the polynomial estimate is used alone since its weighting never drops to zero. The weighted estimate using the spatial and timewise interpolation techniques is:-

$$E_{a,t}=b\left(\frac{p_{a,t}}{\Delta t}+\alpha\left(A_n+A_s\right)+\beta\left(A_e+A_w\right)\right)$$

where $p_{a,t}$ = Estimate based on Chebyshev polynomial

A_n, A_s, A_e, A_w are anomalies for adjacent areas to the north, south, east and west respectively and Δt= length of gap in time (in seasons) between the last observed anomaly and the next observed anomaly.

$\alpha = \frac{1}{4}$ but if A_n or $A_s = 0$, then $\alpha = 0$

$\beta = \frac{1}{4}$ but if A_e or $A_w = 0$, then $\beta = 0$

and

$$b = \frac{1}{\frac{1}{\Delta t} + \alpha + \beta}$$

7.2.2 Spatial and Temporal Interpolation for data on an Equal Area Grid

The irregularity of this grid required that changes be made to the scheme in 7.2.1. Correlations between adjacent 'equal area' areas, usually calculated using all available pairs of data within the eigenanalysis period, are used to help with the interpolation. Only data for neighbouring areas (within 2000Km) which correlate strongly with data for the area ($r > 0.6$) are used in the interpolation process. This is done as follows:

1. Observed data for the area with missing data to be filled within the eigenvector analysis period (the host area) is correlated over all observed seasons with the time series for all neighbouring areas. If $r > 0.6$ the neighbouring area is specified as a reference area.

2. A weighted mean of all reference area mean anomalies is calculated for those areas where data is available and forms the spatial estimate, S. The weights are the correlations between the reference area and the host area. If no reference area data is available, no spatial interpolation is done and the gap is filled using timewise Chebyshev interpolation only.

3. The finally interpolated estimate, E, is a weighted mean of the Chebyshev timewise interpolation estimate, T, and the spatial estimate, S

$$E = rS + (1-r)T$$

where r is a weight defined as follows. If only one reference area is used, r is the correlation between the reference area and host area series. If two reference areas are used, r is the highest of the correlations between host and each reference station + 0.07. If three or more reference stations are used r is the highest of the correlations between the host and the reference stations + 0.1. These numbers are added to allow for the fact that the multiple correlation between the host area and the reference areas is higher than that from any one area. It was decided not to carry out a full analysis each time estimates were made because of computer time limitations.

8.ANALYSIS PERIODS

The periods over which the eigenvectors were calculated have varied. Pre-1990 SST analyses were usually made over the period 1901-80 and occasionally 1951-80. The choice of a 1901 start was based partly on data availability and reliability and partly the need to fully sample the climate changes of the twentieth century (Bottomley et al). Data after 1981 were left mainly to allow an even number of decades to be chosen and partly to allow some independent data. A further problem concerns the ratio of the number of cases sampled in time (320 seasons in 1901-80) and the number of areas or "variables" sampled spatially (294 in 1901-80 10°x10° SST eigenvectors). Thus the latter had a ratio of cases to variables just over unity, marginally acceptable for a reasonably stable analysis, but less than the ideal ratio of 2:1. Because of the low cases to variables ratio, 1951-80 eigenvectors are considerably less reliable statistically than 1901-80 eigenvectors and should be viewed with caution. However real data over the 1951-80 period are much more complete and the analyses were surprisingly in tune with those for 1901-80 though a "climate change" EOF was not found. Such an eigenvector is the leading eigenvector over 1901-80. Although this is not surprising because of the relatively weaker climate change signal over 1951-80, some confounding of climate change with other signals occurs in these EOFs.

With the advent of MOHSST4, improved data became available. The equal area grid, with area sizes equal or equivalent to 10° lat. x 15° lon. areas at the equator, was also devised to reduce the number of variables (184 in MOHSST4 equal area EOFs). As a result, the cases to variables ratio was close to 2:1 in 1900-89 EOFs calculated with these data and was also improved for the 1950-89 EOFs. However the influence of climate change over the period 1950-89 is stronger than over 1951-80 but is still not explicitly resolved in the 1950-89 EOFs. This caused leading 1950-89 EOFs to be more severely corrupted than the 1951-80 EOFs.

All EOFs so far described were calculated using data from all 4 seasons. The effect of confining the data used to particular three month seasons needed to be understood. So separate 10°x10° EOF analyses were made for 1901-80 using data for individual seasons (Folland and Colman, 1988). Additional analyses were also made using equal area grid data over 1900-89 and for annual mean anomalies despite the low cases to variables ratio of the latter. Separate analyses were also made changing the definition of a three month season. The three definitions are: (a) Mar-May, June-Aug etc. (b) April-June, July-Sept etc (c) May-July, Aug-Oct etc.) The effect of such changes on a given type of analysis was found to be minimal. MOHSST3 SST eigenvectors are mostly based on (a) while MOHSST4 and recently made MOHMAT eigenvectors are based on (b).

9.FILTERING OUT DIFFERENT TIMESCALES IN ORIGINAL DATA FOR EOFS BASED ON MOHSST5 DATA

9.1 Use of Kalman Filter

Equal area EOFs of 360 seasonal SST anomalies from 1901-90 based on MOHSST5 show variability over a wide variety of timescales. A useful distinction can be made between interannual to several years variability and inter-decadal climate variability. This division, for example, divides variability connected with ENSO (3-7 years) and variability associated with long term trends. A Kalman filter (Murphy and Katz, 1985, Ng and Young 1990) has been applied to the original SST data after interpolation in order to look at patterns on these two timescales individually, the dividing line being a period of about 10 years. (Other divisions may be made in future work).

The Kalman filter was chosen for several reasons. Firstly it is able to provide a steep boundary between the lower and higher frequency parts of the data. Secondly it is designed in principle to filter non-stationary time series. In a non-stationary series like the global SST series, the non-stationarity can appear as high frequency variability. A term in the Kalman equation, the noise-variance ratio NVR, governs the cut-off period or frequency in the presence of non-stationarity. The measure of non-stationarity at any point in time is assumed to be the linear slope of the data for the local period (here of length near 10 years,i.e. just greater than the cut-off period). Variability for periods greater than this is modelled by a recursive procedure that allows for the continual change of slope as a function of time. NVR is related to the cut-off period by the equation;

$$NVR = 1605\ (f^{4}).$$

For our 90 year (360 season) SST analyses NVR was set to 0.0006 corresponding to a cut off period of 10 years. The exact filter function is shown in Figure 4.

The Kalman filter was separately applied to the SST anomaly time series for each area in the equal area grid. It has been applied to both MOHSST4 and MOHSST5 data. Eigenvectors were made using the low and high frequency components in the same way as with ordinary data. MOHSST5 filtered data eigenvectors were calculated over the period 1901-90 and 1951-90. The filtered data was calculated from 3 month seasonal mean MOHSST5 anomalies over the period 1901-90.

9.2 Use of Leading EOFs to remove trend

A simpler technique of detrending has also been used. Climate change can be removed from a dataset by subtracting the variance of a 'climate change' eigenvector from the raw SST data. The time series of the first EOF for both 1901-80 and 1901-90 are very highly correlated (r=0.95) with global mean SST on the seasonal timescale over the past century. This technique removes the non-linear trend due to climate change (variability represented by other eigenvectors can in principle be removed in the same way). This procedure is given by:

$$D_{t,a} = S_{t,a} - Z_t E_a$$

for season t and area a

$D_{t,a}$ = Detrended SST anomaly for area a and season t
$S_{t,a}$ = Raw SST anomaly for area a and season t
Z_t = Time coefficient for season t
E_a = Eigenvector weight for area a

Climate change eigenvectors that have been used for this purpose are MOHSST3 global $10^{\circ}x10^{\circ}$ covariance eigenvector 1 and MOHSST5 equal area grid correlation eigenvector 1 (Fig. 2). These have similar spatial patterns and time series. The procedure is not exactly the same, in principle, as using a filter that allows for non-linearity in the low frequency signal like the Kalman filter. This is because some high frequency variability is included in the first EOF and is also removed when the variance of the EOF is subtracted.

10.THE EIGENANALYSIS

For most studies, sets of unrotated correlation and covariance eigenvectors have been calculated and varimax rotated EOFs have been calculated as well. One idea behind these choices is to use non-rotated EOFs when we believe (from other evidence) that the pattern is truly global or quasi-global and to use rotation when the remaining patterns appear to be truly regional. The first three unrotated global SST and NMAT eigenvectors over 1901-80 or 1901-90 are reasonably stable no matter what versions of the data or grid type are used or whether correlation or covariance analysis is used (EOF3 does vary somewhat though). The first three EOFs represent global mean SST (EOF1), El Nino pattern (EOF2) and an interhemispheric contrast pattern related to Sahel rainfall (EOF3). Richman (1986) compares the merits of rotated and unrotated analysis but fails to fully bring out the above point which we believe to be crucial to the successful use of EOF analyses of surface marine data.

A number of techniques can be used to produce rotations. So far we have only used the varimax technique. This technique maximizes the sum of the variances of the principal components spatially distributed weights (principal component = eigenvector x square root (variance explained)). This automatically localizes the EOF patterns by concentrating weights one or more regions. Unrotated eigenvectors 4 to 13 have been rotated in most cases; EOFS below about 13 appear to be noise according to standard tests.

The fact that meaningful patterns have been created through eigenanalysis has so far rendered unnecessary the use of more complex factor analysis techniques. Factor analysis techniques are more risky when little is known about the data since the techniques remove a "noise" component. On the other hand if the analysis correctly identifies this noise, the remaining patterns should have increased in their physical meaning. Factor analysis may be tried in future work.

A list of Met Office papers that have used marine EOF analyses is placed in Appendix 1. A full description of the EOF analysis is being prepared for publication with C.K.Folland.

ACKNOWLEDGEMENT

Many thanks to C.K.Folland for numerous discussions and for commenting on the paper.

REFERENCES

Alexander, R.C.
Mobley, R.L.
Monthly average sea surface temperatures and ice-pack limits on a 1° global grid.
Mon.Weather.Rev.,104,143-148 (1976)

Bottomley, M.
Folland, C.K.
Hsiung, J.
Newell, R.
Parker, D.E.
Global Ocean Surface Temperature Atlas.
A Joint project of the UK Meteorological Office and the Massachusetts Institute of Technology,
funded by the UK Dept. of the Environment and the US Navy.
HMSO (1990)

Burden, R.L.
Faires, J.D.
Numerical Analysis (4th ed.).
PWS Kent Publishing company, Boston, USA (1989)

Folland, C.K.
Parker, D.E.
Kates, F.E.
Worldwide marine temperature fluctuations 1856-1981.
Nature, 310, pp 670-673 (1984)

Folland, C.K.
Parker, D.E.
Chadwick, M.
Derivation of Corrections to Sea Surface Temperatures using a model of the Physics of an uninsulated Bucket
Met O 13 Branch Memo 168 (Dec 1986)

Folland, C.K.
Hsiung, J.
Correction of seasonaly-varying biases in uninsulated Bucket sea surface temperature data using a physical model. Met O 13 Branch memo 154 (updated version 1987)

Folland, C.K.
The weighting of data in an EOF analysis
Met O 13 Discussion note 113 (Nov 1988)

Folland C.K.
Colman A.W.
An interim analysis of the leading covariance eigenvectors of worldwide sea surface temperature anomalies for 1901-80.
Met O 13 LRFC 20 (1988)

Folland, C.K.
Parker, D.E
Observed Variations of Sea Surface Temperature.
Proc. Nato Advanced workshop on Climate-Ocean Interaction Oxford, Sept 1988, Kluwer (1990)

Folland, C.K.
Models of the cooling of sea temperature buckets used to correct historical sea surface temperature data.
Met O CTRN 14 (1991)

Minhinick, J.M
Folland, C.K.
The Meteorological Office Historical Sea-Surface Temperature Dataset. Met O 13 Branch memo 137 (1982)

Murphy, A.H.
Katz, R.W.
Probability, Statistics and Decision making in the Atmospheric Sciences. Westview Press, Boulder, Col, USA. (1985)

Ng, C.N.
Young, P.C.
Recursive Estimation and Forecasting of Non-stationary Time Series. J. Forecasting, Vol 9, pp 173-204 (1990)

Parker, D.E.
Folland, C.K.
Global Ocean Surface Temperature Atlas: a presentation.
Proc. NOAA 15th Annual Climate Diagnostics Workshop, Asheville. (1990)

Parker, D.E.
Folland, C.K.
Worldwide surface temperature trends since the mid-19th Century. In 'Greenhouse-gas induced climate change: A critical appraisal of simulations and observations.
M.E. Schlesinger, Ed: Elsevier, Amsterdam, pp 173-193 (1991)

Richman, M.M,
Rotation of principal components.
J.Climatology., 6, pp 293-335 (1986)

Shearman, R.J.
The Meteorological Office Marine Data Bank.
The Marine Observer Vol 53 No. 282, pp 208-217 (1983)

Young, P.C.
Ng,C.N.
Lane, K
Parker, D.
Recursive Forecasting, Smoothing and Seasonal Adjustment of non-stationary Environmental Data.
J. Forecasting, Vol 10, pp 57-89 (1991)

APPENDIX 1

Published or written papers that refer to MOHSST/MOHMAT eigenvector analyses.

Folland C.K, Ward M.N, Colman A.W. — Sahel rainfall in 1986 - an issued forecast & its evaluation. Letter to Weather, 42, pp258-259. (1987)

Ward M.N, Parker D.E, Folland C.K, — Sea surface temperature anomaly patterns & prediction of seasonal rainfall in the Sahel region of Africa. Recent Climate Change - a Regional Approach. Ed: S. Gregory, Belhaven Press, Chapter 15, pp166-178. (1988)

Ward M.N, Brooks S, Folland C.K. — Predictability of seasonal rainfall in the N Nordeste region of Brazil. Ibid, Chapter 21, pp237-251. (1988)

Folland C.K, Parker D.E, Ward M.N, Colman A.W, — Sahel rainfall, N Hemisphere circulation anomalies & worldwide sea temperature changes. Study Week on "Persistent meteo-oceanographic anomalies & teleconnections", Pontifical Academy of Sciences, Rome, 23-27 Sept 1986, Ed: C. Chagas & G Puppi. Pontificiae Academiae Scientiarum Scripta Varia, 69, pp393-436. (1988)

Folland C.K., Ward M.N, Parker D.E, — A multivariate long-range forecast technique for rainfall over the Sahel incorporating eigenvectors of worldwide sea surface temperature. Ibid, pp301-312. (1988)

Folland C.K. — The weighting of data in an EOF analysis. Met O13 Discussion Note 113.

Parker D.E, Folland C.K., Owen J, — The influence of global sea surface temperature on African rainfall. Proc. 1st Tech. Conf. on Met. Research in East & South Africa, Nairobi, 6-9 Jan 1987, Kenya Met Dept. Tech. Conf. No 1/87, pp46-49. (1989)

Folland C.K, Parker D.E, Ward M.N, Jones D.E. — Experiments in the prediction of African rainfall. ibid. pp50-52 (1989)

Ward M.N, Folland C.K, Parker D.E. — Statistical prediction of seasonal rainfall in the Sahel & other tropical regions. Ibid, pp82-86. (1989)

Ward M.N, Folland C.K, Maskell K, Owen J. Rowell D.P. — Forecasting Sahel rainfall - an update. Weather, 45, 122-125. Special Issue for WMO Day 1990 & UN International Decade for Natural Disaster Reduction. (1990)

Ward M.N, Folland C.K, Maskell K, Rowell D.P, Colman A.W, — Understanding & predicting seasonal rainfall in Sub-Saharan Africa. Proc. Third WMO Symposium on Meteorological Aspects of Tropical Droughts with the emphasis on Long-range Forecasting, Niamey, Niger, 30 April-4 May 1990, pp157-162. TMRPR 36 WMO/TD 353. (1990)

Folland C.K., Parker D.E. — Observed variations of sea surface temperature. In: "Climate-Ocean Interaction" NATO/CEC ARW, Oxford, England, Sept 26-30, 1988, pp21-52. Ed: M Schlesinger, Kluwer Academic Publishers. (1990)

IPCC — Section 7 "Observed Climate Variability & Change" in "Climate Change", WMO/UNEP Intergovernmental Panel on Climate Change Scientific Assessment. Cambridge University Press. (1990)

Folland C.K., Parker D.E., Colman A.W., Bevan A.C. — Low frequency variability of the oceans. Extended Abstract submitted for second week of the WMO Training Workshop on Diagnosis & Prediction of Monthly & Seasonal Atmospheric Variations, Nanjing, China, 15-19 Oct 1990, LRFRR 13, p39. WMO/TD 363 (1991)

Folland C.K., Owen J Ward M.N., Colman A.W. — Prediction of seasonal rainfall in the Sahel region using empirical & dynamical methods. J. Forecasting, 10, pp21-56. (1991)

Folland C.K., Rowell D.P., Ward M.N., Maskell K, — Predictability & Physical Factors Influencing Sahel Rainfall Variation. ICTP/WMO International Technical Conference on Long-range Forecasting Research, Trieste, April 8-12 1991, LRFRR 14, pp83-87. WMO/TD 395 (1991)

Ward M.N., Folland C.K, — Empirical prediction of NE Brazil (North Nordeste) Rainfall. Ibid, pp297-300. (1991)

Parker D.E, Folland C.K, — Surface temperature trends since the mid nineteenth century. Proc. US Dept. of Energy Workshop on Greenhouse Gas-Induced Climate Change, Amherst, Mass, 8-12 May 1989. Ed: M Schlesinger, Kluwer Academic Press. (1991)

Ward M.N, Folland C.K, — Statistical prediction of N E Brazil seasonal rainfall rainfall from sea surface temperatures. Int. J. Climatol., (1991)

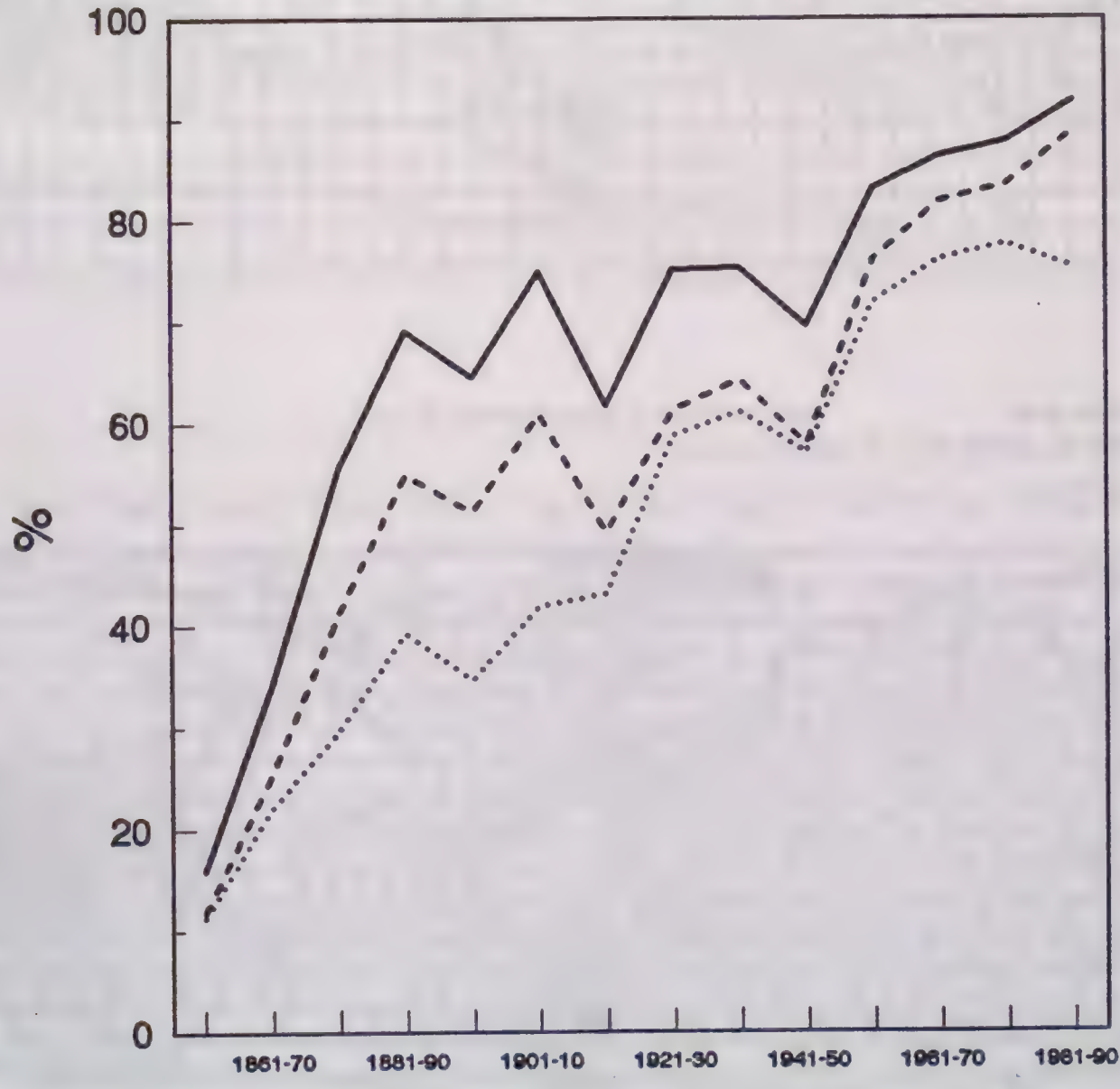
Ward,M.N. — Cross Spectral Analysis and Coherence Significance Levels -Idealised examples and an application to Sahel Rainfall, Hadley Centre Internal Technical Note No. 8. (1991)

Ward M.N, Owen J.A Folland C.K, Farmer G. — The relationship between sea surface temperature anomalies & summer rainfall in Africa 4-20°N. Report of the Conference of Commonwealth Meteorologists, 19-22 June 1989, Shinfield, pp 86-92. Also in LRFC 32 (1989)

FIGURE 1

Coverage of global Ocean by MOHSST data at 5 degree square resolution

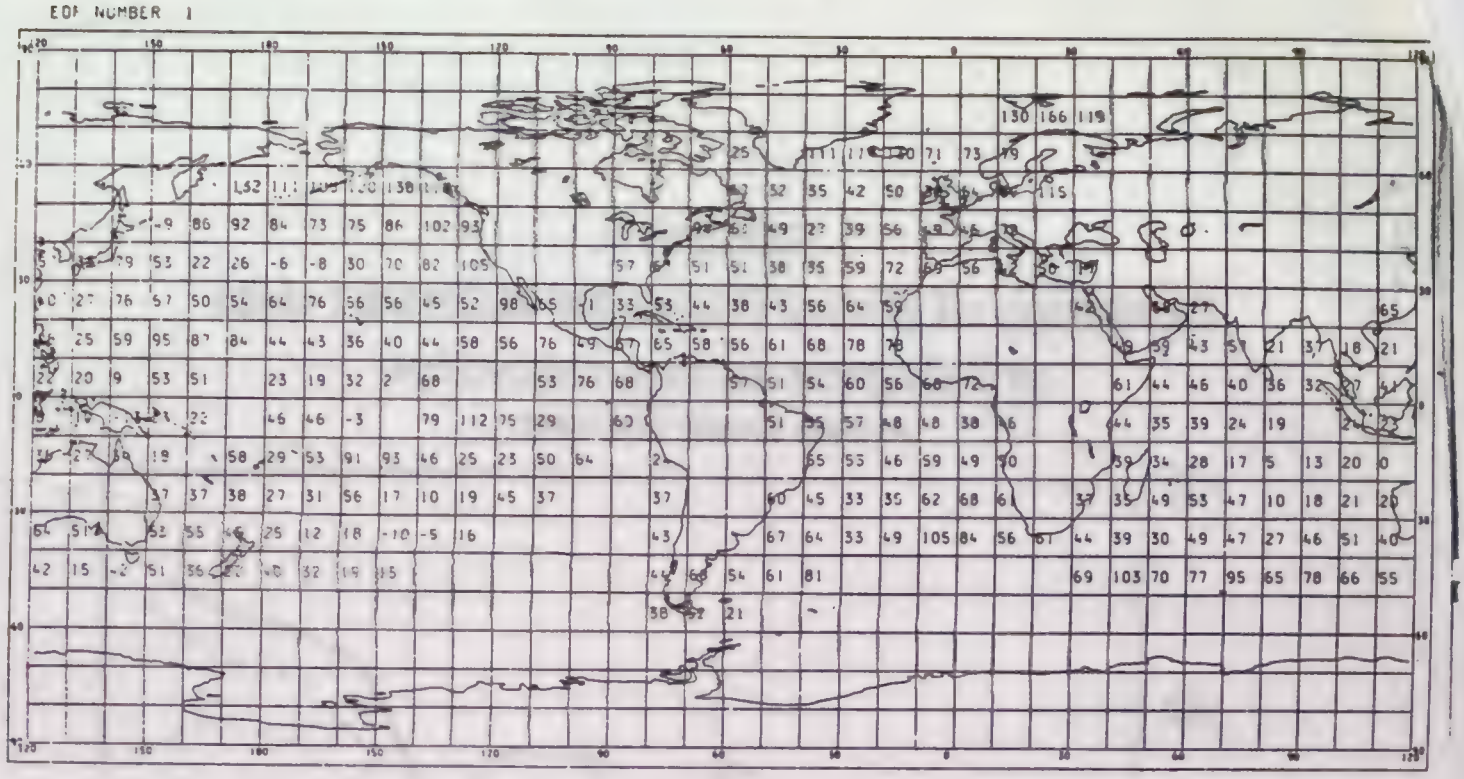
for each decade 1851-1990



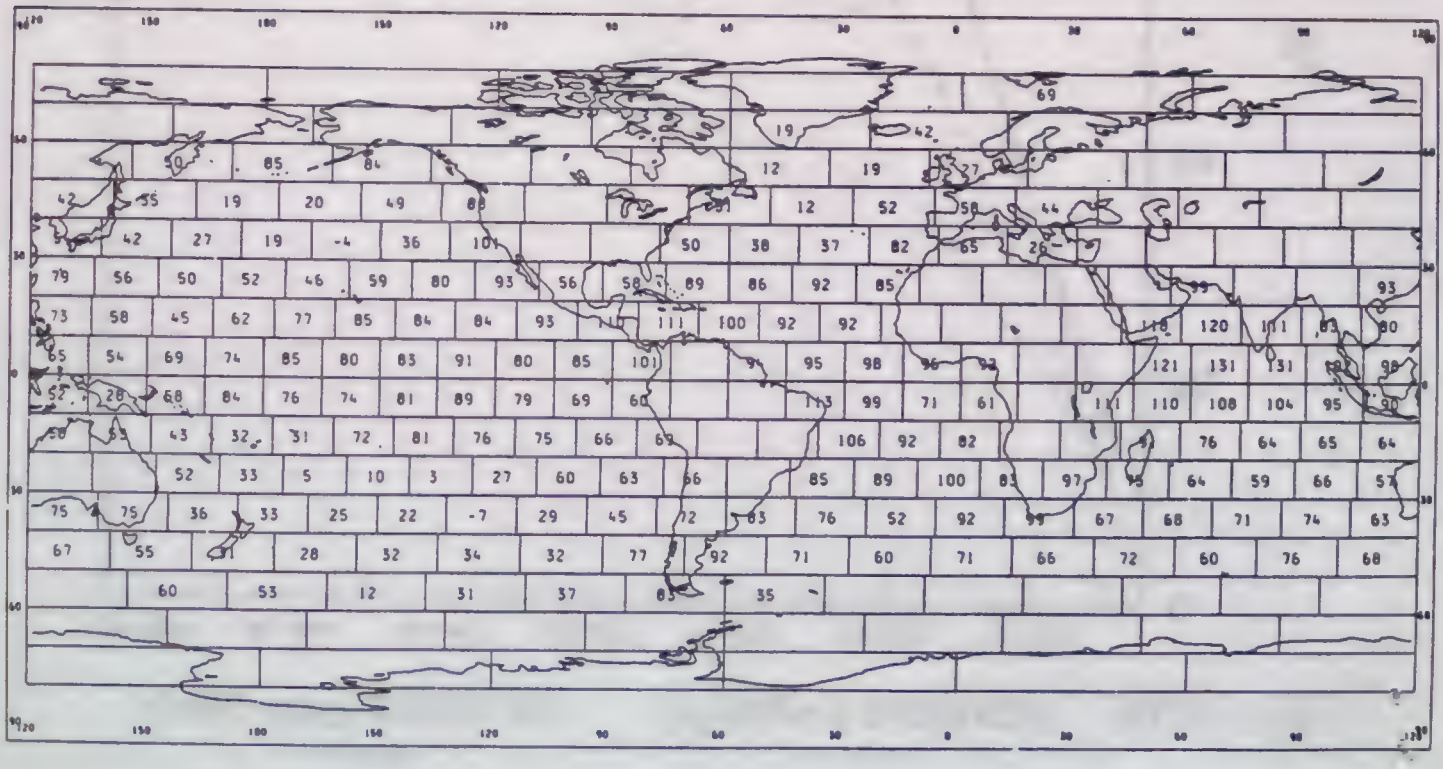
MOHSST3 ..... MOHSST5 (no fillings) - - - - - MOHSST5 (gaps filled) ———

Coverage is assumed to be complete if there is an SST anomaly for every 5x5 degree square for every 3 month season.

a)



b)



IPCC CORRECTIONS ADDED

FIGURE 2 a) Leading covariance eigenvector using MOHSST3 data and 10°x10° grid and b) leading correlation eigenvector using MOHSST5 data and equal area grid.

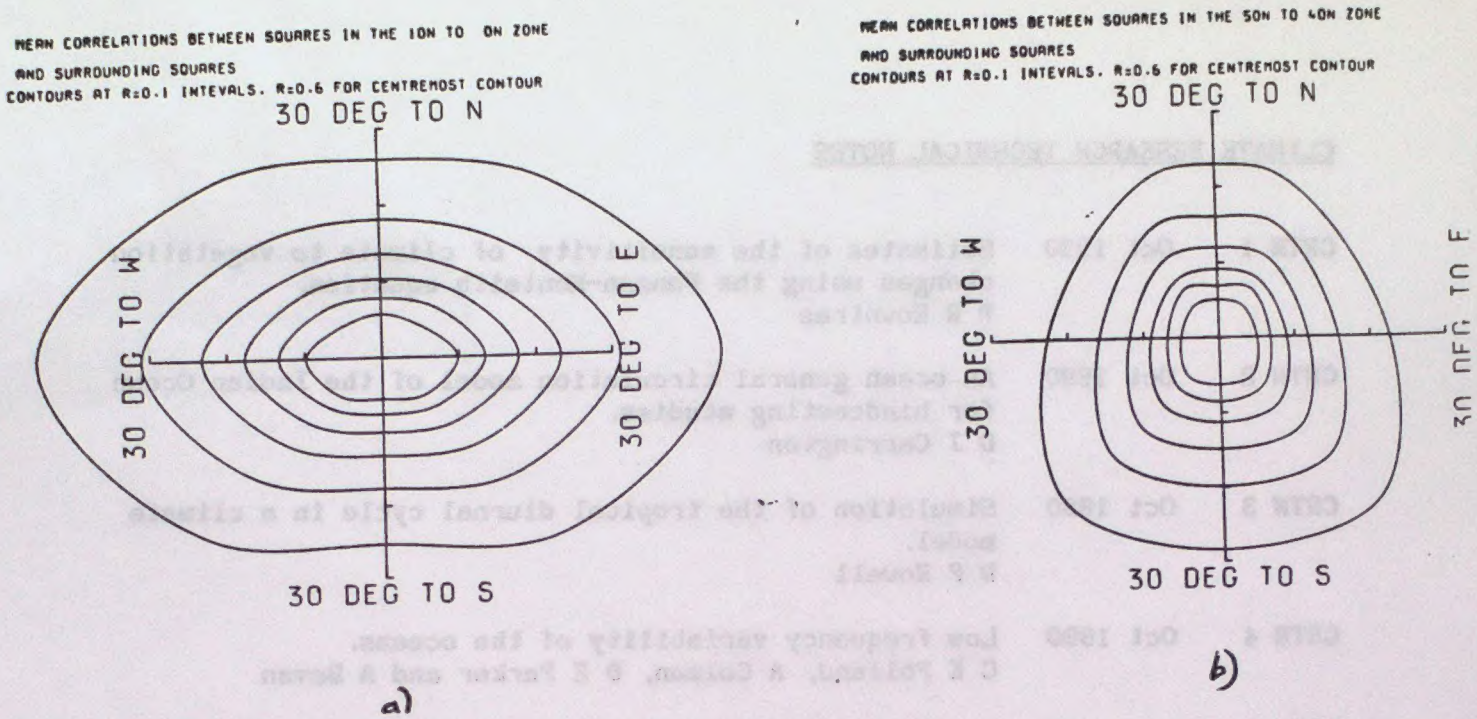


FIGURE 3 Mean correlation function for 10^0x10^0 square anomalies between a) 0^0 and 10^0N and b) 40^0 and 50^0N.

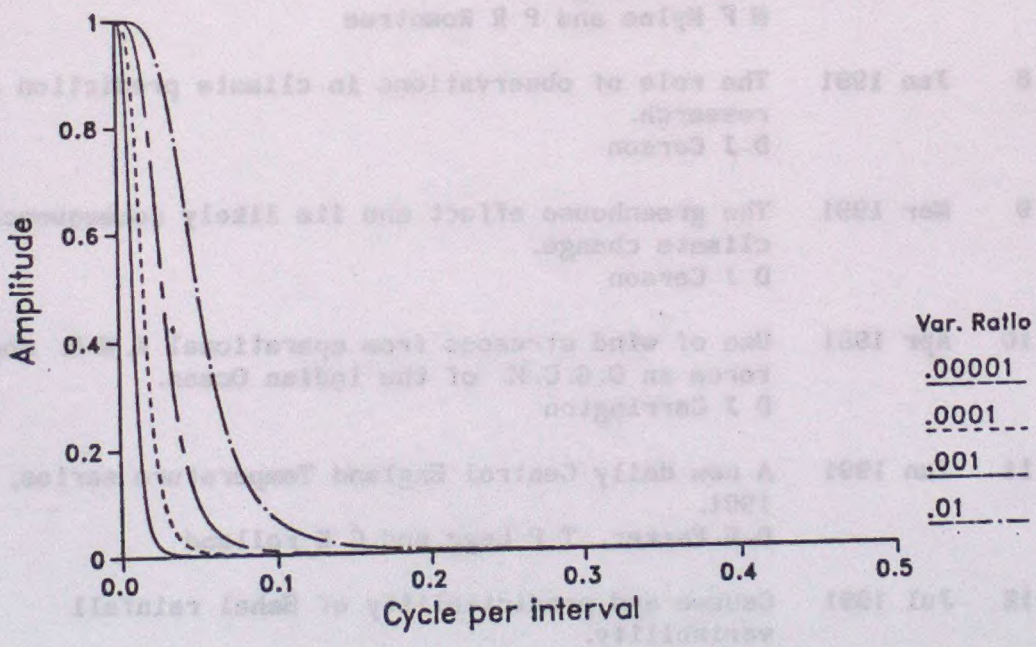


FIGURE 4 Frequency response of the Kalman Filter. For filtering SST data, Var Ratio = 0.0006.

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 1	Oct 1990	Estimates of the sensitivity of climate to vegetation changes using the Penman-Monteith equation. P R Rowntree
CRTN 2	Oct 1990	An ocean general circulation model of the Indian Ocean for hindcasting studies. D J Carrington
CRTN 3	Oct 1990	Simulation of the tropical diurnal cycle in a climate model. D P Rowell
CRTN 4	Oct 1990	Low frequency variability of the oceans. C K Folland, A Colman, D E Parker and A Bevan
CRTN 5	Dec 1990	A comparison of 11-level General Circulation Model Simulations with observations in the East Sahel. K Maskell
CRTN 6	Dec 1990	Climate Change Prediction. J F B Mitchell and Qing-cun Zeng
CRTN 7	Jan 1991	Deforestation of Amazonia - modelling the effects of albedo change. M F Mylne and P R Rowntree
CRTN 8	Jan 1991	The role of observations in climate prediction and research. D J Carson
CRTN 9	Mar 1991	The greenhouse effect and its likely consequences for climate change. D J Carson
CRTN 10	Apr 1991	Use of wind stresses from operational N.W.P. models to force an O.G.C.M. of the Indian Ocean. D J Carrington
CRTN 11	Jun 1991	A new daily Central England Temperature series, 1772-1991. D E Parker, T P Legg and C K Folland
CRTN 12	Jul 1991	Causes and predictability of Sahel rainfall variability. D P Rowell, C K Folland, K Maskell, J A Owen, M N Ward
CRTN 13	Jul 1991	Modelling changes in climate due to enhanced CO_2, the role of atmospheric dynamics, cloud and moisture. C A Senior, J F B Mitchell, H Le Treut and Z-X Li

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 14	Sep 1991	Sea temperature bucket models used to correct historical SST data in the Meteorological Office. C K Folland
CRTN 15	Aug 1991	Modelling climate change, and some potential effects on agriculture in the U.K. P R Rowntree, B A Callander and J Cochrane
CRTN 16	Aug 1991	The Boreal Forests and Climate G Thomas and P R Rowntree
CRTN 17	Aug 1991	Development of a Stratosphere-Troposphere Data Assimilation System. R Swinbank.
CRTN 18	Sep 1991	A study of asynchronous coupling using a simple climate model. M K Davey.
CRTN 19	Sep 1991	The Oceanic Carbon Cycle. N K Taylor
CRTN 20	Nov 1991	Worldwide ocean-atmosphere surface fields in Sahel wet and dry years using provisionally corrected surface wind data. M N Ward
CRTN 21	Dec 1991	Coupled tropical ocean global atmosphere models at the UKMO. M K Davey, C Gordon, S Ineson and S Lawrence
CRTN 22	Dec 1991	Empirical parameterisation of tropical ocean-atmosphere coupling. M Allen, M K Davey, D L T Anderson and P D Killworth
CRTN 23	Feb 1992	The temporal evolution of Equatorial Currents in the Indian Ocean D L T Anderson, D J Carrington, R A Corry and C Gordon
CRTN 24	Feb 1992	Stratospheric analyses provided by the U.K. Meteorological Office M J Bailey, A O'Neill and V D Pope
CRTN 25	Feb 1992	Modelling interannual variability in the Indian Ocean using momentum fluxes from the UKMO and ECMWF operational weather analyses. D L T Anderson and D J Carrington

CLIMATE RESEARCH TECHNICAL NOTES

CRTN 26	Mar 1992	A GCM simulation of the impact of Amazonian deforestation on climate using an improved canopy representation J Lean and P R Rowntree
CRTN 27	Mar 1992	The parameterization of rainfall interception in GCMs A J Dolman and D Gregory
CRTN 28	Jun 1992	Development of worldwide marine data eigenvectors since 1985 A W Colman